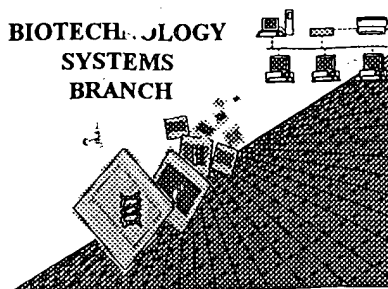


**RAW SEQUENCE LISTING**  
**ERROR REPORT**

BIOTECHNOLOGY  
SYSTEMS  
BRANCH



0590  
1005

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following computer readable form:

Application Serial Number: 09/891,064

Source: OIPK

Date Processed by STIC: 8/17/2001

THE ATTACHED PRINTOUT EXPLAINS DETECTED ERRORS.

PLEASE FORWARD THIS INFORMATION TO THE APPLICANT BY EITHER:

- 1) INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE APPLICANT, WITH A NOTICE TO COMPLY or,
- 2) TELEPHONING APPLICANT AND FAXING A COPY OF THIS PRINTOUT, WITH A NOTICE TO COMPLY

FOR CRF SUBMISSION QUESTIONS, PLEASE CONTACT MARK SPENCER, 703-308-4212.

FOR SEQUENCE RULES INTERPRETATION, PLEASE CONTACT ROBERT WAX, 703-308-4216.

PATENTIN 2.1 e-mail help: [patin21help@uspto.gov](mailto:patin21help@uspto.gov) or phone 703-306-4119 (R. Wax)

PATENTIN 3.0 e-mail help: [patin3help@uspto.gov](mailto:patin3help@uspto.gov) or phone 703-306-4119 (R. Wax)

TO REDUCE ERRORED SEQUENCE LISTINGS, PLEASE USE THE CHECKER VERSION 3.0 PROGRAM, ACCESSIBLE THROUGH THE U.S. PATENT AND TRADEMARK OFFICE WEBSITE. SEE BELOW:

### Checker Version 3.0

The Checker Version 3.0 application is a state-of-the-art Windows based software program employing a logical and intuitive user-interface to check whether a sequence listing is in compliance with format and content rules. Checker Version 3.0 works for sequence listings generated for the original version of 37 CFR §§1.821 – 1.825 effective October 1, 1990 (old rules) and the revised version (new rules) effective July 1, 1998 as well as World Intellectual Property Organization (WIPO) Standard ST.25.

Checker Version 3.0 replaces the previous DOS-based version of Checker, and is Y2K-compliant. Checker allows public users to check sequence listings in Computer Readable form (CRF) before submitting them to the United States Patent and Trademark Office (USPTO). Use of Checker prior to filing the sequence listing is expected to result in fewer errored sequence listings, thus saving time and money.

Checker Version 3.0 can be down loaded from the USPTO website at the following address:

<http://www.uspto.gov/web/offices/pac/checker>

## Raw Sequence Listing Error Summary

### ERROR DETECTED      SUGGESTED CORRECTION

SERIAL NUMBER: 09/891,064

ATTN: NEW RULES CASES: PLEASE DISREGARD ENGLISH "ALPHA" HEADERS, WHICH WERE INSERTED BY PTO SOFTWARE

- 1 ☐ Wrapped Nucleics      The number/text at the end of each line "wrapped" down to the next line. This may occur if your file was retrieved in a word processor after creating it. Please adjust your right margin to .3; this will prevent "wrapping."
- 2 ☐ Invalid Line Length      The rules require that a line not exceed 72 characters in length. This includes white spaces.
- 3 ☒ Misaligned Amino Numbering      The numbering under each 5\* amino acid is misaligned. Do not use tab codes between numbers; use space characters, instead.
- 4 ☐ Non-ASCII      The submitted file was not saved in ASCII(DOS) text, as required by the Sequence Rules. Please ensure your subsequent submission is saved in ASCII text.
- 5 ☐ Variable Length      Sequence(s) \_\_\_\_\_ contain n's or Xaa's representing more than one residue. Per Sequence Rules, each n or Xaa can only represent a single residue. Please present the maximum number of each residue having variable length and indicate in the <220>-<223> section that some may be missing.
- 6 ☐ PatentIn 2.0 "bug"      A "bug" in PatentIn version 2.0 has caused the <220>-<223> section to be missing from amino acid sequence(s) \_\_\_\_\_. Normally, PatentIn would automatically generate this section from the previously coded nucleic acid sequence. Please manually copy the relevant <220>-<223> section to the subsequent amino acid sequence. This applies to the mandatory <220>-<223> sections for Artificial or Unknown sequences.
- 7 ☐ Skipped Sequences (OLD RULES)      Sequence(s) \_\_\_\_\_ missing. If intentional, please insert the following lines for each skipped sequence:  
 (2) INFORMATION FOR SEQ ID NO:X: (insert SEQ ID NO where "X" is shown)  
 (i) SEQUENCE CHARACTERISTICS: (Do not insert any subheadings under this heading)  
 (xi) SEQUENCE DESCRIPTION:SEQ ID NO:X: (insert SEQ ID NO where "X" is shown)  
 This sequence is intentionally skipped  
  
 Please also adjust the "(ii) NUMBER OF SEQUENCES:" response to include the skipped sequences.
- 8 ☐ Skipped Sequences (NEW RULES)      Sequence(s) \_\_\_\_\_ missing. If intentional, please insert the following lines for each skipped sequence.  
 <210> sequence id number  
 <400> sequence id number  
 000
- 9 ☐ Use of n's or Xaa's (NEW RULES)      Use of n's and/or Xaa's have been detected in the Sequence Listing.  
 Per 1.823 of Sequence Rules, use of <220>-<223> is MANDATORY if n's or Xaa's are present.  
 In <220> to <223> section, please explain location of n or Xaa, and which residue n or Xaa represents.
- 10 ☐ Invalid <213> Response      Per 1.823 of Sequence Rules, the only valid <213> responses are: Unknown, Artificial Sequence, or scientific name (Genus/species). <220>-<223> section is required when <213> response is Unknown or is Artificial Sequence
- 11 ☐ Use of <220>      Sequence(s) \_\_\_\_\_ missing the <220> "Feature" and associated numeric identifiers and responses.  
 Use of <220> to <223> is MANDATORY if <213> "Organism" response is "Artificial Sequence" or "Unknown." Please explain source of genetic material in <220> to <223> section.  
 (See "Federal Register," 06/01/1998, Vol. 63, No. 104, pp. 29631-32) (Sec. 1.823 of Sequence Rules)
- 12 ☐ PatentIn 2.0 "bug"      Please do not use "Copy to Disk" function of PatentIn version 2.0. This causes a corrupted file, resulting in missing mandatory numeric identifiers and responses (as indicated on raw sequence listing). Instead, please use "File Manager" or any other manual means to copy file to floppy disk.

# RAW SEQUENCE LISTING

## PATENT APPLICATION US/09/891,064

DATE: 08/11/2001  
TIME: 02:33:31

INPUT SET: S36597.raw

*use all upper-case letters for headings*

This Raw Listing contains the General Information Section and up to the first 5 pages.

*pp 2-5*

1 GENERAL INFORMATION: SEQUENCE LISTING

2  
3 (1) General Information

Does Not Comply  
Corrected Diskette Needed

4  
5 (i) APPLICANTS: James M. Anderson  
6 Christina M. Van Itallie

7  
8 (ii) TITLE OF INVENTION: Human Occludin, Its Uses  
9 and Enhancement of Drug Absorption Using Occlu-  
10 din Inhibitors

11  
12 (iii) NUMBER OF SEQUENCES: 6

13  
14 (iv) CORRESPONDENCE ADDRESS: *insert colon after all headings*

15  
16 (A) ADDRESSEE: Yale University Medical School  
17 Section of Pulmonary and Critical Care Medicine  
18 Department of Internal Medicine

19 (B) STREET: 333 Cedar Street, LCI 105

20 (C) CITY: New Haven

21 (D) STATE: Connecticut

22 (E) COUNTRY: United States of America

23 (F) ZIP CODE: 065220-8057

24  
25 (v) COMPUTER READABLE FORM: *←*

26  
27 (A) MEDIUM TYPE: 3.5" 1.44 Mb diskette

28 (B) COMPUTER: IBM PC

29 (C) OPERATING SYSTEM: MS DOS

30 (D) SOFTWARE: Word Processing

31  
32 (vi) CURRENT APPLICATION DATA: *←*

33  
34 (A) APPLICATION NUMBER:

35 (B) FILING DATE: June 25, 2001

36 (C) CLASSIFICATION:

37  
38 (vii) PRIOR APPLICATION DATA: *←*

39  
40 (A) APPLICATION NUMBER: U.S. 09/142,732

41 (B) FILING DATE: September 15, 1998

42 ~~(C) CLASSIFICATION: 530-350.000~~

43  
44 (viii) ATTORNEY INFORMATION: ATTORNEY/AGENT INFORMATION:

45  
46 (A) NAME: Mary M. Krinsky

*FYI: all U.S. applications filed on or after July 1, 1998 and which cannot claim a prior application filed before July 1, 1998, need to be in new sequence listing format.*

RAW SEQUENCE LISTING  
PATENT APPLICATION US/09/891,064DATE: 08/11/2001  
TIME: 02:33:32NUMBER  
INPUT SET: S36597.raw

(B) REGISTRATION NO.: 32423

(C) REFERENCE/DOCKET NUMBER: OCR-754.CIP

*Delete all  
hard page  
break codes*

## (ix) TELECOMMUNICATION INFORMATION

(A) TELEPHONE NUMBER: 203-773-9544

(B) TELEFAX NUMBER: 203-773-1183 (2) INFORMATION FOR SE

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2312

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE:

(A) DESCRIPTION: cDNA

## (v) FRAGMENT TYPE: complete sequence

## (ix) FEATURE:

(A) NAME/KEY: human occludin

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

65	GCCTCTCTCC	ATCAGACACC	CCAAGGTTCC	ATCCGAAGCA	GGCGGAGCAC	50
66	CGAACGCACC	CCGGGGTGGT	CAGGGACCCC	CATCCGTGCT	GCCCCCTAGG	100
67	AGCCCCGCGC	TCTCCTCTGC	GCCCCGCCTC	TCGGGCCGCA	ACATCGCGCG	150
68	GTTCCCTTAA	CAGCGCGCTG	GCAGGGTGTG	GGAAGCAGGA	CCGCGTCCTC	200
69	CCGCCCCCTC	CCATCCGAGT	TTCAGGTGAA	TTGGTCACCG	AGGGAGGAGG	250
70	CCGACACACC	ACACCTACAC	TCCC GCGTCC	ACCTCTCCCT	CCCTGCTTCC	300
71	TCTTGGCGGA	GGCGGCAGGA	ACCGAGAGCC	AGGTCCAGAG	CGCCGAGGAG	350
72	CCGGTCTAGG	ACGCAGCAGA	TTGGTTTATC	TTGGAAGCTA	AAGGGCATTG	400
73	CTCATCCTGA	AGATCAGCTG	ACCATTGACA	ATCAGCCATG	TCATCCAGGC	450
74	CTCTTGAAAG	TCCACCTCCT	TACAGGCCTG	ATGAATTCAA	ACCGAATCAT	500
75	TATGCACCAA	GCAATGACAT	ATATGGTGGA	GAGATGCATG	TTCGACCAAT	550
76	GCTCTCTCAG	CCAGCCTACT	CTTTTTACCC	AGAAGATGAA	ATTCTTCACT	600
77	TCTACAAATG	GACCTCTCCT	CCAGGATGTA	TTCGGATCCT	GTCTATGCTC	650
78	ATTATTGTGA	TGTGCATTGC	CATCTTTGCC	TGTGTGGCCT	CCACGCTTGC	700
79	CTGGGACAGA	GGCTATGGAA	CTTCCCTTTT	AGGAGGTAGT	GTAGGCTACC	750
80	CTTATGGAGG	AAGTGGCTTT	GGTAGCTACG	GAAGTGGCTA	TGGCTATGGC	800
81	TATGGTTATG	GCTATGGCTA	CGGAGGCTAT	ACAGACCCAA	GAGCAGCAAA	850
82	GGGCTTCATG	TTGGCCATGG	CTGCCTTTTG	TTTCATTGCC	GCGTTGGTGA	900
83	TCTTTGTTAC	CAGTGTTATA	AGATCTGAAA	TGTCCAGAAC	AAGAAGATAC	950
84	TACTTAAGTG	TGATAATAGT	GAGTGCTATC	CTGGGCATCA	TGGTGTTTAT	1000
85	TGCCACAATT	GTCTATATAA	TGGGAGTGAA	CCCAACTGCT	CAGTCTTCTG	1050
86	GATCTCTATA	TGGTTCACAA	ATATATGCCC	TCTGCAACCA	ATTTTATACA	1100
87	CCTGCAGCTA	CTGGACTCTA	CGTGGATCAG	TATTTGTATC	ACTACTGTGT	1150
88	TGTGGATCCC	CAGGAGGCCA	TTGCCATTGT	ACTGGGGTTC	ATGATTATTG	1200
89	TGGCTTTTGC	TTTAATAAAT	TTCTTTTGCTG	TGAAAACTCG	AAGAAAGATG	1250
90	GACAGGTATG	ACAAGTCCAA	TATTTTGTGG	GACAAGGAAC	ACATTTATGA	1300
91	TGAGCAGCCC	CCCAATGTCG	AGGAGTGGGT	TAAAAATGTG	TCTGCAGGCA	1350
92	CACAGGACGT	GCCTTCACCC	CCATCTGACT	ATGTGGAAAG	AGTTGACAGT	1400
93	CCCATGGCAT	ACTCTTCCAA	TGGCAAAGTG	AATGACAAGC	GGTTTTATCC	1450
94	AGAGTCTTCC	TATAAATCCA	CGCCGGTTCC	TGAAGTGGTT	CAGGAGCTTC	1500
95	CATTAACCTC	GCCTGTGGAT	GACTTCAGGC	AGCCTCGTTA	CAGCAGCGGT	1550
96	GGTAACTTTG	AGACACCTTC	AAAAAGAGCA	CCTGCAAAGG	GAAGAGCAGG	1600
97	AAGGTCAAAG	AGAACAGAGC	AAGATCACTA	TGAGACAGAC	TACACAACCTG	1650
98	GCGGCGAGTC	CTGTGATGAG	CTGGAGGAGG	ACTGGATCAG	GGAATATCCA	1700
99	CCTATCACTT	CAGATCAACA	AAGACAACCTG	TACAAGAGGA	ATTTTGTACAC	1750

# RAW SEQUENCE LISTING PATENT APPLICATION US/09/891,064

DATE: 08/11/2001  
TIME: 02:33:32

INPUT SET: S36597.raw

```

100 TGGCCTACAG GAATACAAGA GCTTACAATC AGAACTTGAT GAGATCAATA 1800
101 AAGAACTCTC CCGTTTGGAT AAAGAATTGG ATGACTATAG AGAAGAAAGT 1850
102 GAAGAGTACA TGGCTGCTGC TGATGAATAC AATAGACTGA AGCAAGTGAA 1900
103 GGGATCTGCA GATTACAAAA GTAAGAAGAA TCATTGCAAG CAGTTAAAGA 1950
104 GCAAATTGTC ACACATCAAG AAGATGGTTG GAGACTATGA TAGACAGAAA 2000
105 ACATAGAAGG CTGATGCCAA GTTGTTTGAG AAATTAAGTA TCTGACATCT 2050
106 CTGCAATCTT CTCAGAAGGC AAATGACTTT GGACCATAAC CCCGGAAGCC 2100
107 AAACCTCTGT GAGCATCACA AAGTTTTGGG TTGCTTTAAC ATCATCAGTA 2150
108 TTGAAGCATT TTATAAATCG CTTTGTGATA TCAACTGGGC TGAACAACCTC 2200
109 CAATTAAGGA TTTTATGCTT TAAACATTGG TTCTTGATT AAGAATGAAA 2250
110 TACTGTTTGA GGTTTTAAAG CCTTAAAGGA AGGTTCTGGT GTGAACTAAA 2300
111 CTTTCACACC CC 2312

```

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 522
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE:

(A) DESCRIPTION: polypeptide

(v) FRAGMENT TYPE: complete sequence

(ix) FEATURE:

(A) NAME/KEY: human occludin

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Ser Ser Arg Pro Leu Glu Ser Pro Pro Pro Tyr Arg Pro Asp  
5 10 15

Glu Phe Lys Pro Asn His Tyr Ala Pro Ser Asn Asp Ile Tyr Gly  
20 25 30

Gly Glu Met His Val Arg Pro Met Leu Ser Gln Pro Ala Tyr Ser  
35 40 45

Phe Tyr Pro Glu Asp Glu Ile Leu His Phe Tyr Lys Trp Thr Ser  
50 55 60

Pro Pro Gly Val Ile Arg Ile Leu Ser Met Leu Ile Ile Val Met  
65 70 75

Cys Ile Ala Ile Phe Ala Cys Val Ala Ser Thr Leu Ala Trp Asp  
80 85 90

Arg Gly Tyr Gly Thr Ser Leu Leu Gly Gly Ser Val Gly Tyr Pro  
95 100 105

Tyr Gly Gly Ser Gly Phe Gly Ser Tyr Gly Ser Gly Tyr Gly Tyr  
110 115 120

Gly Tyr Gly Tyr Gly Tyr Gly Tyr Gly Tyr Thr Asp Pro Arg  
125 130 135

*do not use  
TABS*

*Misaligned amino acid numbers - see  
item 3 on Error Summary Sheet*

**RAW SEQUENCE LISTING**  
PATENT APPLICATION *US/09/891,064*

DATE: 08/11/2001  
TIME: 02:33:33

**INPUT SET: S36597.raw**

[illegible]

Same error

# RAW SEQUENCE LISTING PATENT APPLICATION US/09/891,064

DATE: 08/11/2001  
TIME: 02:33:33

INPUT SET: S36597.raw

206		395		400		405
207						
208	Gly Glu Ser Cys Asp	Glu Leu Glu Glu Asp Trp Ile Arg Glu Tyr				
209		410		415		420
210						
211	Pro Pro Ile Thr Ser Asp Gln Gln Arg Gln Leu Tyr Lys Arg Asn					
212		425		430		435
213						
214	Phe Asp Thr Gly Leu Gln Glu Tyr Lys Ser Leu Gln Ser Glu Leu					
215		440		445		450
216						
217	Asp Glu Ile Asn Lys Glu Leu Ser Arg Leu Asp Lys Glu Leu Asp					
218		455		460		465
219						
220	Asp Tyr Arg Glu Glu Ser Glu Glu Tyr Met Ala Ala Ala Asp Glu					
221		470		475		480
222						
223	Tyr Asn Arg Leu Lys Gln Val Lys Gly Ser Ala Asp Tyr Lys Ser					
224		485		490		495
225						
226	Lys Lys Asn His Cys Lys Gln Leu Lys Ser Lys Leu Ser His Ile					
227		500		505		510
228						
229	Lys Lys Met Val Gly Asp Tyr Asp Arg Gln Lys Thr					
230		515		520		
231						
232						

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE:

(A) DESCRIPTION: peptide

(v) FRAGMENT TYPE: synthetic peptide

(ix) FEATURE:

(D) OTHER INFORMATION: construct used in experiments

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

247	Cys Asp Arg Gly Tyr Gly Thr Ser Leu Leu Gly Gly Ser Val Gly
248	5 10 15
249	
250	Tyr Pro Tyr Gly Gly Ser Gly Phe Gly
251	20
252	
253	
254	

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24
- (B) TYPE: amino acid

**SEQUENCE VERIFICATION REPORT**  
**PATENT APPLICATION US/09/891,064**DATE: 08/11/2001  
TIME: 02:33:33**INPUT SET: S36597.raw**

Line	Error	Original Text
12	Number of Sequences (6) Doesn't Equal Actual Count (5)	(iii) NUMBER OF SEQUENCES: 6
42	Wrong Classification	(C) CLASSIFICATION: 530-350.000
54	Unknown or Misplaced Identifier	(i) SEQUENCE CHARACTERISTICS:
55	Unknown or Misplaced Identifier	(A) LENGTH: 2312
56	Unknown or Misplaced Identifier	(B) TYPE: nucleic acid
57	Unknown or Misplaced Identifier	(C) STRANDEDNESS: single
58	Unknown or Misplaced Identifier	(D) TOPOLOGY: linear
59	Unknown or Misplaced Identifier	(ii) MOLECULE TYPE:
60	Unknown or Misplaced Identifier	(A) DESCRIPTION: cDNA
61	Unknown or Misplaced Identifier	(v) FRAGMENT TYPE: complete sequence
62	Unknown or Misplaced Identifier	(ix) FEATURE:
63	Unknown or Misplaced Identifier	(A) NAME/KEY: human occludin
64	Unknown or Misplaced Identifier	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1: